ENSIAME

Optimisation:

De l'estimation paramétrique à l'apprentissage, une ballade entre théorie et pratique

S. Delprat

Chapitre 4 – Dérivation numérique

1) Méthode des pénalités

2) Programmation Quadratique Sequentielle *(SQP Sequential Quadratic Programming)*

- *Avec uniquement des contraintes égalités*
- *Avec des contraintes égalités & inégalités*

Dérivation numérique

En pratique, le critère à optimiser n'est pas nécessairement une fonction « simple ». Son évaluation peut nécessiter un ou plusieurs simulation.

Dans ce cas, on ne connait pas la valeur de son gradient ou Hessien

=> Il faut pouvoir évaluer numériquement les dérivées (Gradient, Jacobienne, Hessien, etc.)

Dérivation numérique

Cas scalaire: $f \in C^1 : \mathbb{R} \to \mathbb{R}$

$$\frac{df(x)}{dx} = \lim_{\varepsilon \to 0^+} \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

Dérivée à gauche

$$\frac{\widehat{df(x)}}{dx} = \frac{f(x) - f(x-\varepsilon)}{\varepsilon}$$

1 évaluation de f

$$\left| \frac{\widehat{df(x)}}{dx} - \frac{df(x)}{dx} \right| = O(\varepsilon)$$

Dérivée centrée

$$\frac{\widehat{df(x)}}{dx} = \frac{f(x+\varepsilon) - f(x-\varepsilon)}{2\varepsilon}$$

2 évaluations de f

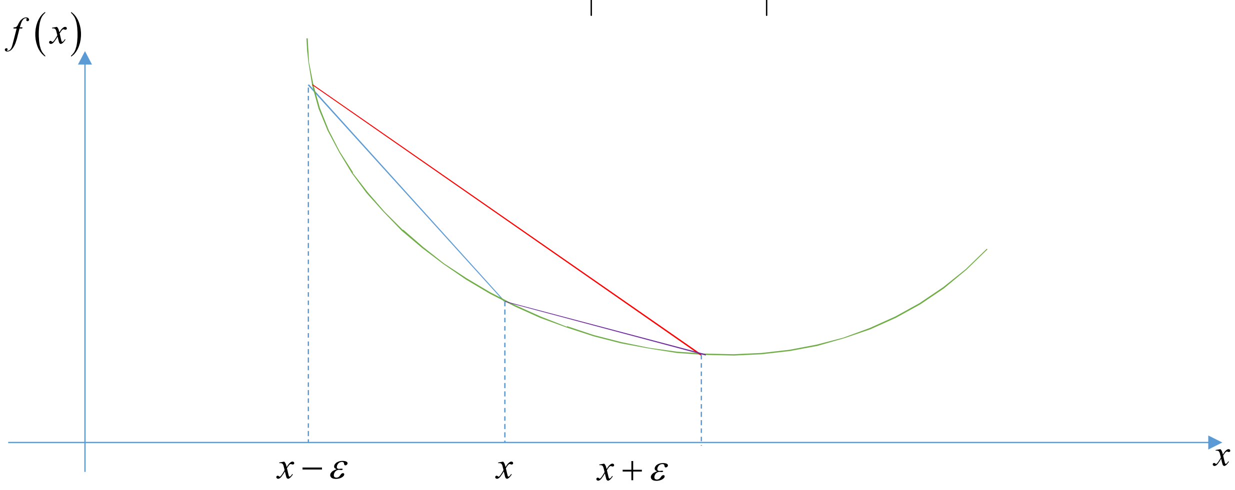
$$\left| \frac{\widehat{df(x)}}{dx} - \frac{df(x)}{dx} \right| = O(\varepsilon^2)$$

Dérivée à droite

$$\frac{\widehat{df(x)}}{dx} = \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

1 évaluation de f

$$\left| \frac{\widehat{df(x)}}{dx} - \frac{df(x)}{dx} \right| = O(\varepsilon)$$



Dérivation numérique

<u>Cas scalaire:</u> $f \in C^1 : \mathbb{R} \to \mathbb{R}$

Dérivée à gauche

$$\widehat{\frac{df(x)}{dx}} = \frac{f(x) - f(x-\varepsilon)}{\varepsilon}$$

Dérivée centrée

$$\widehat{\frac{df(x)}{dx}} = \frac{f(x+\varepsilon) - f(x-\varepsilon)}{2\varepsilon}$$

Dérivée à droite

$$\widehat{\frac{df(x)}{dx}} = \frac{f(x+\varepsilon) - f(x)}{\varepsilon}$$

<u>Choix de ε:</u>

- Trop petit : pb numérique
- Trop grand : erreur d'estimation

u : Floating point relative accuracy

Double 64bits	Single 32bits
u=2.2204x10^{-16}	u=1.1921x10^{-7}

Sous certaines hypothèses :	Dérivée à droite et à gauche	Dérivée centrée
Choix de ε:	$\varepsilon = \sqrt{u}$	$\varepsilon = u^{1/3}$
Erreur lié au codage des nombres	$\sqrt{u}$	$u^{2/3}$
Erreur d'estimation de la dérivée	$O(\varepsilon)$	$O(\varepsilon^2)$

Dérivation numérique

<u>Cas multivariable:</u> $f \in C^1 : \mathbb{R}^{nx} \to \mathbb{R}$

Le gradient est une collection de dérivées scalaires $\widehat{\nabla f(x)} \approx (g_i)$

	Dérivée à gauche	Dérivée centrée	Dérivée à droite
g_i	$g_i = \dfrac{f(x) - f(x - \varepsilon_i)}{\varepsilon}$	$g_i = \dfrac{f(x + \varepsilon_i) - f(x - \varepsilon_i)}{2\varepsilon}$	$g_i = \dfrac{f(x + \varepsilon_i) - f(x)}{\varepsilon}$
Nombre d'évaluations de f	$nx + 1$	$2nx$	$nx + 1$

Vecteur de perturbation

$$\varepsilon_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \varepsilon \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$ i^{ème} ligne

Différentiation automatique – *Complex step differentiation*

Cas scalaire: $f \in C^{+\infty} : \mathbb{R} \to \mathbb{R}$

Formule de Taylor suivant l'axe imaginaire

$$f(x+i\varepsilon) \approx f(x) + i\varepsilon \frac{df(x)}{dx} - \frac{\varepsilon^2}{2!}\frac{d^2 f(x)}{dx^2} - i\frac{\varepsilon^3}{3!}\frac{d^3 f(x)}{dx^3} \ldots$$

Si on néglige les termes de plus haut niveaux

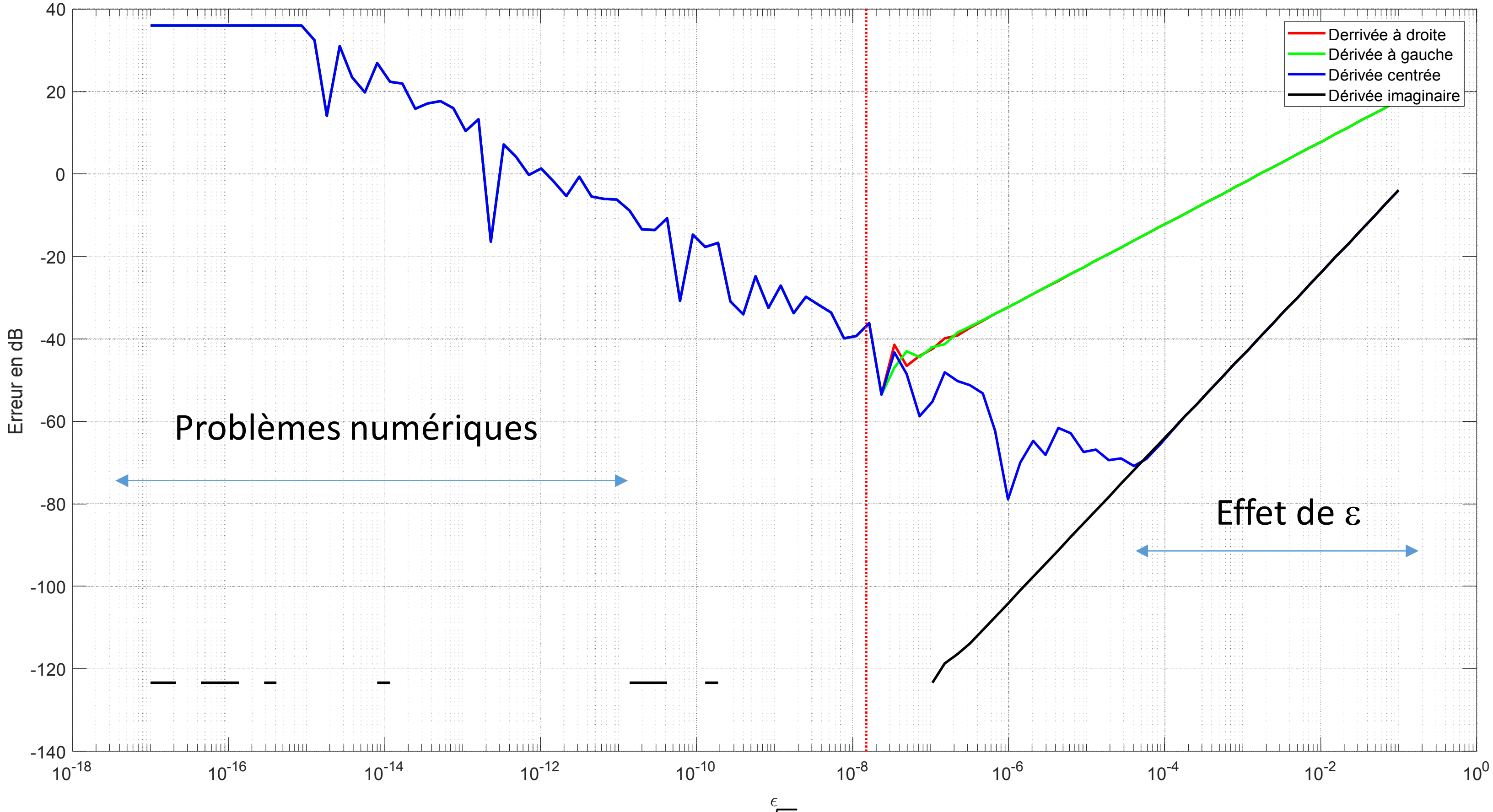
$$f(x+i\varepsilon) \approx f(x) + i\varepsilon \frac{df(x)}{dx} + O(\varepsilon^2)$$

$$\frac{df(x)}{dx} = \frac{im(f(x+i\varepsilon))}{\varepsilon} + O(\varepsilon^2)$$

=> Conclusion si on dispose d'une bibliothèque de calcul qui gère les imaginaires (e.g. Matlab), on obtient une très bonne estimation numérique à l'ordre 2 de la dérivée en 1 évaluation de f !!

Différentiation automatique

$f\left(x^4\right) \quad x = 10$



$\varepsilon = \sqrt{u} = 2.2204 \cdot 10^{-16}$

Différentiation automatique – dual numbers

Complexes : $c=a+ib \qquad i^2=-1$

Dual : $d=a+\varepsilon b \qquad \varepsilon^2=0$

Partie réelle   Partie dualle

$$f(x+h\varepsilon)\approx f(x)+h\varepsilon\frac{df(x)}{dx}+\underbrace{h^2\frac{\varepsilon^2}{2!}\frac{d^2f(x)}{dx^2}+h\frac{\varepsilon^3}{3!}\frac{d^3f(x)}{dx^3}\ldots}_{=0 \text{ car } \varepsilon^2=0}$$

$$f(x+h\varepsilon)=f(x)+h\varepsilon\frac{df(x)}{dx} \quad\Longrightarrow\quad \sin(x+h\varepsilon)=\sin(x)+\cos(x)h\varepsilon$$

$$f(x+1\varepsilon)=f(x)+\varepsilon\frac{df(x)}{dx} \quad\Longrightarrow\quad \frac{df(x)}{dx}=dual\big(f(x+1\varepsilon)\big)$$

$\Rightarrow$ Nécessite une bibliothèque de calcul sur les nombres duaux (simple en C++)

$\Rightarrow$ Obtention de la valeur exacte de la dérivée

Hessien

En général, on n'a pas besoin de l'estimation du Hessien, mais de l'inverse du Hessien.
Les formules de CFGS et DFP permettent d'estimer les 2. Ne pas se tromper.

BFGS: $$H_k = H_{k-1} + \frac{yy^T}{y^T s} - \frac{H_{k-1} s s^T H_{k-1}}{s^T H_{k-1} s}$$ $$s = x^k - x^{k-1}, y = \nabla f\left(x^k\right) \nabla f\left(x^{k-1}\right)$$

$$\left(H_k\right)^{-1} = \left(I - \frac{sy^T}{y^T s}\right)\left(H_{k-1}\right)^{-1}\left(I - \frac{sy^T}{y^T s}\right) + \frac{ss^T}{y^T s}$$

Propriété : si $H_0 > 0$ alors $H_k > 0$ pour tout k